4956 44

Reliability and Validity of Chest Radiograph Surveillance Programs* ADMINISTRATIVE RECORD

Christina C. Lawson, PhD; Megan K. LeMasters, MD;
Grace Kawas Lemasters, PhD; Susan Simpson Reutman, PhD;
Carol H. Rice, PhD; and James E. Lockey, MD, FCCP

Study objectives: Due to the lack of consensus in the literature in the use of posteroanterior (PA) vs PA with right and left oblique views as the optimum radiograph surveillance methodology to investigate pleural changes, a study was undertaken to evaluate the reliability, sensitivity, and specificity of these two approaches.

Design: Three experienced radiologist B readers used the 1980 International Labor Office classification system for pneumoconiosis to independently read chest radiographs of workers with individual identifiers masked. All radiographs were read first as a PA view only. Unknown to the B readers, each subject's PA was then matched to his or her corresponding right and left oblique views (film triad) and re-read several weeks later.

Setting and participants: The respiratory health of 652 workers exposed to refractory ceramic fiber was assessed as part of cross-sectional and longitudinal surveillance programs.

Measurements and results: κ Statistics for interreader and intrareader reliability between the PA view and film triad methods were calculated. Sensitivity, specificity, and positive predictive value were assessed by comparing the initial cross-sectional study to the longitudinal study. The film triad method had considerably higher interreader reliability (κ = 0.59) compared to the PA-only method (κ = 0.44). Results from the initial cross-sectional study were then compared to findings evaluated longitudinally. The film triad again was superior, demonstrating a positive predictive value of 73.7% compared to only 47.8% for the PA method.

Conclusions: It is our recommendation that the film triad method be used in surveillance studies where both parenchymal and pleural changes are anticipated. *(CHEST 2001; 120:64–68)*

Key words: pleura; precision; radiographs; reliability; sensitivity; specificity; surveillance

Abbreviations: HRCT = high-resolution CT; PA = posteroanterior; RCF = refractory ceramic fiber

Abnormalities of the pleura are a common manifestation of asbestos exposure.[1] More recently, occupational exposure to refractory ceramic fiber (RCF) has been shown to be associated with pleural plaques.[2,3] Pleural plaques are areas of thickening of the parietal pleura most commonly located at the midcostal area and posterior costal area and at the dome of the diaphragm. The presence of pleural calcification is also described. Pleural plaques are considered a marker of past exposure.[1,4-7]

Posteroanterior (PA) chest radiographs are gener-

ally used in pulmonary surveillance programs for the detection of pleural changes seen with asbestos exposure. The extent of a pleural plaque can be difficult to determine when using this view alone, however, and pleural abnormalities can be confused with fat deposition, bone shadows, and extrathoracic muscle.[8,9] The addition of a left and right 45° oblique view (hereafter referred to as a *film triad*) may increase sensitivity by including lung surface not seen tangentially in the PA view.[10,11]

Variability in the interpretation of chest radiographs has long been recognized as a serious limitation in their use. Though some of the sources of variability can be controlled, such as film quality and training of the reader, others are inherent to the reading of chest radiographs and include age and weight of the subjects, location of the plaques, and type of abnormality.[12-15] The interpretation of each radiograph is dependent on the observer. Measurement quality can be assessed through indexes of reliability (precision), validity (accu-

racy), sensitivity, and specificity. If reliability (agreement) among readers is low, then the usefulness of the chest radiograph as a surveillance tool is questionable. Improving the interreader reliability by controlling for sources of variability is, therefore, extremely important in occupational health surveillance studies.

Due to a lack of consensus in the use of PA or film triads, and the large variability in the reliability of previous studies, the current study had two goals: (1) to evaluate the percent detection and the intrareader and interreader agreement using PA views alone vs the film triad for identifying pleural plaques, and (2) to determine the sensitivity and specificity of these two methodologies by comparing the findings of the positive changes identified during an initial cross-sectional evaluation with the results after multiple follow-up visits in a 12-year longitudinal study.

MATERIALS AND METHODS

Chest radiographs for the present study were obtained as part of a respiratory medical surveillance program that included both former and current employees at two plant locations, one in New York and one in Indiana, involved with the manufacture of RCF and RCF products.[2,3] The study was approved by the Institutional Review Board at the University of Cincinnati Medical Center; informed consent was obtained from all subjects. Several medical facilities in each local community were asked to submit chest radiographs for evaluation before one was selected as a provider. After the chest radiographs were obtained, they were sent to the University of Cincinnati Medical Center and then masked of all identifying information. Approximately 25% of the study radiographs were "control" radiographs from individuals without a known fiber exposure obtained from the same medical facilities.

All the chest radiographs were interpreted independently by the same three radiologist B readers for both the cross-sectional and the longitudinal portion of the study using the 1980 International Labor Office classification of radiographs for pneumoconioses.[16] These B readers were certified by the National Institute for Occupational Safety and Health and averaged approximately 8 years of experience in B reading at the outset of the study. Film quality was scored from 1 (good) through 4 (unreadable) by all B readers, with their median score used for determination of film quality. Radiographs rated 1, 2, or 3 were considered acceptable. Independent agreement by at least two B readers (the median reading) was used for recording whether or not a radiograph demonstrated a pleural change.

Pleural plaques vary in size and may calcify over time. Pleural abnormalities were defined by site and type. The site definitions included chest wall, diaphragm, and costophrenic angle. In evaluating type, the width and extent of the thickening, and whether it was diffuse or circumscribed (plaque) in the chest wall was noted. Costophrenic angle-blunting alone was not considered a pleural change in this study.

The PA radiographs were evaluated first. Then, several weeks later, the worker's PA radiograph was matched with his or her two oblique radiographs and re-read as a film triad by all three B readers without knowledge of their previous reading. The B readers were unaware that they were engaging in a reliability study and that they were reading each subject's radiographs twice. Readers were blinded to all identifying information about the subjects' demographic, medical, or occupational histories and to other radiologists' interpretations.

A κ statistic was calculated to measure interreader agreement for pleural changes among the three readers, correcting for the degree of agreement that is expected by chance alone.[17,18] Interreader reliabilities within either the PA or film triad viewing methods were determined in pairwise reader comparisons (such as reader 1 vs reader 2 for PA readings). Intrareader agreement was examined by comparing each reader's PA radiograph interpretation to his film triad interpretation and by comparing agreement of two out of the three readers between methods. The κ values of the following ranges were interpreted: < 0.40 equals poor agreement, between 0.40 and 0.75 is moderate to good, and between 0.75 and 1.0 is excellent agreement.[19] Agreement that deviated from chance was tested by using the χ^2 statistic. Significant differences between intrareader and interreader reliabilities were tested using the "best test" of Dunn and Clark[20] for comparison of correlation coefficients, based on Fisher's z transformation. Due to the relatively low prevalence of pleural changes (3.1%), the proportion of specific agreement was used in addition to κ.[18,21] The proportion of specific agreement quantity is the conditional probability that the second reader will also make an assignment to the same category as a randomly selected first reader.

There were 652 current and former employees who had at least 1 year of employment and who participated in the initial cross-sectional evaluation.[3] After the initial evaluation, the subjects in this study were followed for approximately 12 years, receiving radiographic examinations (film triads) approximately every 3 years. The longitudinal analysis included 75.5% of the original cohort (n = 492) who had both a minimum of 5 years of latency since date of hire in a production job and who also obtained two or more film triad evaluations. Of the 492 workers, 65, 149, 229, 38, 10, and 1 worker(s) were evaluated two, three, four, five, six, or seven times, respectively. Therefore, this longitudinal study included approximately 1,750 film triad readings. Once a subject had a positive reading by two of three readers, then the reading of that first positive radiograph plus all subsequent readings was taken to define case status for this longitudinal study.

RESULTS

For the initial cross-sectional testing period, > 96% of the radiographs had a median quality rating of 1 or 2, and no radiographs were rated unreadable (score of 4). In the longitudinal study, approximately 90% of the radiographs had a film quality rating of 1 or 2, and < 1% had a median reading of 4.

Table 1 describes pleural changes from the initial cross-sectional study, comparing each reader, the median reading (two of three B readers), and each method (PA only and the film triad). Ninety-five percent (19 of 20) of pleural changes were plaques and will hereafter be referred to as plaques. Overall, the film triad method had a 44% decrease in positive radiograph readings compared with the PA-only views, 9 vs 5, respectively (Table 1). For the median reading, 15 radiographs were scored as positive for plaques by two of three readers on both the PA view and the film triad (Table 1). The PA-only method had κ values ranging from 0.40 to 0.47, with an

Table 1—*Pleural Plaque Cases, Detection Rates, and Reader Agreement Statistics*[*]

Variables	Cases (+) PA	Cases (+) Film Triad	Cases (+) PA (+) Triad	Cases (+) PA (−) Triad	Cases (−) PA (+) Triad	Change From PA to Triad[†]	Intrareader κ Values
Reader 1	33	20	13	20	7	− 65%	0.47
Reader 2	16	19	12	4	7	+ 75%	0.68
Reader 3	46	34	24	22	10	− 55%	0.57
Median (two of three readers)	24	20	15	9	5	− 44%	0.67
Interreader κ values	0.44	0.59					

[*]Data are presented as No. (−) = negative; (+) = positive.
[†]Percentage change determined as (column 5 − column 4)/column 5 × 100.

overall κ of 0.44, and the film triad method had higher κs ranging from 0.50 to 0.67, with an overall κ of 0.59. The film triad method indicated an improvement of approximately 34%.

Next, we examined intrareader reliabilities by comparing each reader with themselves across the two methods (eg, reader 1 PA view vs reader 1 film triad). The percentage of change of the detection rates by reader from their PA-only reading compared to their film triad method is also shown in Table 1. When comparing positive readings on the PA view and film triad methods for readers 1, 2, and 3, their intrareader reliability κ values were 0.47, 0.68, and 0.57, respectively. The κ for the median reading (two of three readers) was 0.67. Reader 2 had the lowest plaque detection rates, the highest between-method reliability, and was the only reader with an increase in plaque detection from the PA-only method to the film triad method. Because a small proportion of the study population had positive readings, supplemental proportion of specific agreement values were calculated[18] and were comparable for both interreader and intrareader reliabilities. When comparing overall interreader reliabilities, the PA-only view had a lower reliability (κ = 0.44) than the film triad (κ = 0.59), and this difference was statistically significant (p value < 0.001).

Using the results of the longitudinal follow-up study as the "gold standard," we then evaluated the sensitivity, specificity, and positive predictive values of the cross sectional analysis of the PA view vs film triad methods. As shown in Table 2, in a comparison of positive radiographs from the cross-sectional study to the longitudinal study, the sensitivity of the PA radiograph was 64.7%, the specificity was 97.5%, and the positive predictive value was 47.8%. Using the same comparison for the film triad, the sensitivity was higher at 82.4%; the specificity remained high at 98.9%, and the positive predictive value also increased to 73.7%.

DISCUSSION

In the current cross-sectional study, the addition to the PA radiograph of a right and left 45° oblique radiograph (film triad) decreased the number of positive cases from 24 to 20. The interreader reliability increased from the PA radiograph to the film triad from κ = 0.44 to κ = 0.59. The intrareader reliabilities were moderate, with an overall κ of 0.67. When comparing the cross-sectional results to the longitudinal study, the sensitivity, specificity, and positive predictive values were all superior in the film triad method. Based on these calculations, the film triad is a more valid and reliable method than the PA view alone under conditions found in our study.

Table 3 summarizes previous studies of the reliability and validity of both PA radiographs and film triads used for surveillance studies. These previous studies demonstrated wide variations in interreader and intrareader agreement, with κ values ranging from 0.06 to 0.83. In contrast to the current study, Sheers et al,[22] Baker and Greene,[11] and Reger et al[23] demonstrated that the film triad method increased the number of positive findings by 6.2%, 44%, and

Table 2—*Sensitivity and Specificity of the PA and Film Triad Methods*[*]

	Longitudinal Study		
Cross-sectional	Positive (n = 17)	Negative (n = 475)	Total (n = 492)
PA films[†]			
Positive	11	12	23 +
Negative	6	463	469
Film triad[‡]			
Positive	14	5	19 +
Negative	5	470	475

[*]Data are presented as No.
[†]Sensitivity = 11 of 17 (64.7%); specificity = 463 of 475 (97.5%); positive predictive value = 11 of 23 (47.8%).
[‡]Sensitivity = 14 of 17 (82.4%); specificity = 470 of 475 (98.9%); positive predictive value = 14 of 19 (73.7%).

Table 3—*Comparison of Reader Reliability Using PA and Oblique Films in the Published Literature**

Source	Methodology				Descriptive Results		κ Statistics		Assessed Outcome
	Unreadable, † %	Readers‡	Independent Reader	Blind History	Subject and Controls, No.	Detection of PA and FT, %	Interreader PA/FT	Intrareader PA/FT	
Parker et al (1989)[13]	2.3	2 B	Yes	Yes	299		0.58		Any pleural thickening
		2 B	Yes	Yes	553 and C		0.96		Pleural plaque by laterality
		2 B	Yes	Yes	583 and C		0.39		Any pleural change
Bourbeau et al (1990)[9]		2 B	Yes	Yes	182 and 24 C		0.57	0.63 to 0.83	Diaphragm abnormalities
							0.68	0.84 to 0.81	Pleural calcification
							0.15 to 0.44	0.12 to 0.75	Chest wall abnormalities
Musch et al (1985)[15]	0	3 B	Yes		392 and 351 C		0.52 to 0.75		Profusion of combined, small opacity
Fromkin et al (1990)[21]	15.9	6 B	Yes	Yes	421		0.06 to 0.40		Pleural thickening
Sheers et al (1978)[22]	27	5	Yes		1,884				Pleural thickening
Baker and Greene (1982)[11]	1	1 B			100	PA = 43		0.82	Unilateral and bilateral pleural thickening
						FT = 52			
Reger et al (1992)[23]	24.5	3	Yes	Yes	555	PA = 12.8	0.33 PA/0.23 FT	0.53 PA/0.73 FT	Pleural plaque
						FT = 26.0			
Sherman et al (1988)[10]	0	2		Yes	489	PA = 23.9			Pleural or parenchymal abnormalities
						FT = 19.4			

*FT = film triad; C = control radiographs.
†Poor films defined as unreadable.
‡B indicates B reader.
§κ < 0.40 = poor agreement; 0.40 ≤ κ < 0.75 = moderate to good agreement; κ > 0.75 = excellent agreement.[19]

103%, respectively. Sherman et al,[10] however, showed it to decrease by 19%. κ Values also varied with the type of abnormality being examined. Although Reger et al[23] had an increase in detection, the interreader agreement decreased by 30% for the film triad, from κ = 0.33 to κ = 0.23. Only 24.5% (555 of 2,266) of the radiographs in that study had a quality grading of acceptable or excellent.

Film quality is highly subjective and has been found to be inversely related to the difficulty the physician encounters in interpreting the radiograph.[24] The current study had excellent film quality, indicating that with high-quality films, the film triad is preferable over PA view alone. As with the current study, previous studies with better agreement appear to use more experienced B readers, good film quality, and control radiographs.

Another reason for the wide range of agreement in previous studies is the use of too few or newly certified B readers. Musch et al[15] used three independent B readers, film quality grading, and exclusion of unreadable films to achieve high reliability values ranging from 0.52 to 0.75, similar to those found in our study. Ducatman et al[13] also found that, of 23 readers, the "expert" interpreters (National Institute for Occupational Safety and Health B-reader course instructors) diagnosed fewer abnormalities than did other readers.

Though the current study indicates that the film triad method is superior to the PA view alone, both methods suffer from limitations inherent to radiograph technology. Findings on the chest radiograph have been found to be normal on 10 to 20% of patients with asbestosis,[25] and 11% of patients with no occupational asbestos exposure were found to have abnormalities that might have been interpreted as asbestosis.[26] Thus, some cases will go undetected and some will be falsely identified as positive. CT and high-resolution CT (HRCT) are both considered diagnostic "gold standards." CT and to a lesser extent HRCT have higher radiation exposure and financial costs and are traditionally used as follow-up on subjects who had positive clinical findings but normal or unclear radiographs.[27] These increases in radiation exposures and costs are not now regarded as acceptable in a screening program where most individuals are anticipated to be disease free. If the surveillance program included primarily high-risk subjects, however, then CT/HRCT could be considered for its superior accuracy.

CONCLUSION

Given our findings and the stated considerations, it is our recommendation that film triads are the preferred

method for chest radiograph surveillance studies. Diligence must be maintained, however, to ensure high standards for radiographs and selection of only experienced B readers. In addition, the study team must maintain a high rate of participation in follow-up studies in order to minimize bias in the findings.

REFERENCES

1 Hillerdal C. Pleural plaques and risk for bronchial carcinoma and mesothelioma: a prospective study. Chest 1994; 105:144–150

2 Lemasters G, Lockey J, Rice C. Radiographic changes among workers manufacturing refractory ceramic fiber and products. Ann Occup Hyg 1994; 38(suppl 1):745–751

3 Lockey JE, Lemasters G, Rice C, et al. Refractory ceramic fiber exposure and pleural plaques. Am J Respir Crit Care Med 1996; 154:1405–1410

4 Dujic Z, Eterovic D, Tocilj J. Association between asbestos-related pleural plaques and resting hyperventilation. Scand J Work Environ Health 1993; 19:346–351

5 Baker EL, Dagg T, Greene RE. Respiratory illness in the construction trades: I. The significance of asbestos-associated pleural disease among sheet metal workers. J Occup Med 1985; 27:483–489

6 Bourbeau J, Ernst P. Chrome J, et al. The relationship between respiratory impairment and asbestos-related pleural abnormality in an active work force. Am Rev Respir Dis 1990; 142:837–842

7 Weill H, Jones R. The chest roentgenogram as an epidemiologic tool: report of a workshop. Arch Environ Health; 1975; 30:435–439

8 Vix VA. Extrapleural costal fat. Radiology 1974; 112:563–565

9 Collins JD, Pagani JJ. Extrathoracic musculature mimicking pleural lesions. Radiology 1978; 129:21–22

10 Sherman CB, Barnhart S, Rosenstock L. Use of oblique chest roentgenograms in detecting pleural disease in asbestos-exposed workers. J Occup Med 1988; 30:681–683

11 Baker EL, Greene R. Incremental value of oblique chest radiography in the diagnosis of asbestos-induced pleural disease. Am J Ind Med 1982; 3:17–22

12 Parker DL, Bender AP, Hankinson S, et al. Public health implications of the variability of 'B' readings for pleural changes. J Occup Med 1989; 31:775–780

13 Ducatman AM, Yang WN, Forman SA. B-readers and asbestos medical surveillance. J Occup Med 1988; 30:644–647

14 Rossiter CE. Initial repeatability trials of the UICC-Cincinnati classification of the radiographic appearances of pneumoconiosis. Br J Ind Med 1972; 29:407–419

15 Musch DC, Higgins IT, Landis JR. Some factors influencing inter-observer variation in classifying simple pneumoconiosis. Br J Ind Med 1985; 42:346–349

16 International Labor Office. Guidelines for the use of ILO international classification of radiographs of pneumoconioses. Occupational safety and health series, No. 22 (revised). Geneva, Switzerland: International Labor Office, 1980

17 Light RJ. Measures of response agreement for qualifying data: some generalizations and alternatives. Psychol Bull 1971; 76:365–377

18 Fleiss JL. Statistical methods for rates and proportions. New York, NY: Wiley, 1981

19 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159–174

20 Dunn OJ, Clark V. Correlation coefficients measured on the same individuals. J Am Stat Assoc 1969; 65:366–377

21 Hoehler FK. Bias and prevalence effects on κ viewed in terms of sensitivity and specificity. J Clin Epidemiol 2000; 53:499–503

22 Sheers C, Rossiter CE, Gilson JC, et al. UK Naval Dockyards Asbestosis Study: radiologic methods in the surveillance of workers exposed to asbestos. Br J Ind Med 1978; 35:195–203

23 Reger RB, Ames RG, Merchant JA, et al. The detection of thoracic abnormalities using posteroanterior (PA) vs PA and oblique roentgenograms. Chest 1982; 81:290–295

24 Merchant JA, Reger RB. Classification of the chest radiograph for the pneumoconioses. In: Merchant JA, ed. Occupational respiratory disease. Washington DC: U.S. Department of Health and Human Services, 1986; 113–122

25 Rockoff SD, Schwartz A. Roentgenographic underestimation of early asbestosis by International Labor Office classification: analysis of data and probabilities. Chest 1988; 5:1088–1091

26 Epstein DM, Miller WT, Bresnitz EA, et al. Applications of ILO classification to a population without industrial exposure: findings to be differentiated from pneumoconiosis. AJR Am J Roentgenol 1984; 142:53–58

27 Muller NL. Clinical value of high-resolution CT in chronic diffuse lung disease. AJR Am J Roentgenol 1991; 157:1163–1170

28 Frumkin H, Pransky C, Cosmatos I. Radiologic detection of pleural thickening. Am Rev Respir Dis 1990; 142:1325–1330